KLASIFIKASI DAN PEMBANGKITAN INDEKS UNTUK PENCARIAN KOLEKSI DOKUMEN ONLINE DENGAN MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER DAN VECTOR SPACE MODEL

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai gelar Strata Satu
Program Studi Informatika

Disusun oleh:

ARBA SASMOYO

M0511007

PROGRAM STUDI INFORMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS SEBELAS MARET

SURAKARTA

2016

SKRIPSI

KLASIFIKASI DAN PEMBANGKITAN INDEKS UNTUK PENCARIAN KOLEKSI DOKUMEN ONLINE DENGAN MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER DAN VECTOR SPACE MODEL

Disusun oleh:

Arba Sasmoyo

M0511007

Diajukan untuk memenuhi sebagian persyaratan memperoleh gelar Strata Satu

Program Studi Informatika

PROGRAM STUDI INFORMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS SEBELAS MARET

SURAKARTA

2016

SKRIPSI

KLASIFIKASI DAN PEMBANGKITAN INDEKS UNTUK PENCARIAN KOLEKSI DOKUMEN ONLINE DENGAN MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER DAN VECTOR SPACE MODEL

Disusun oleh:

Arba Sasmoyo
M0511007

Skripsi ini telah disetujui untuk dipertahankan dihadapan dewan penguji,

Pada tanggal: 19 Januari 2016

Pembimbing 1,	Pembimbing 2,
Ristu Saptono S.Si., M.T.	Dr. Wiranto M.Kom., M.Cs.
NIP. 197902102002121001	NIP. 196612301993021001

iii

SKRIPSI

KLASIFIKASI DAN PEMBANGKITAN INDEKS UNTUK PENCARIAN KOLEKSI DOKUMEN ONLINE DENGAN MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER DAN VECTOR SPACE MODEL

Disusun oleh:

Arba Sasmoyo

M0511007

Telah dipertahankan dihadapan Dewan Penguji,

pada tanggal: 19 Januari 2016

Susunan Dewan Penguji

1. **Ristu Saptono S.Si., M.T.** ( )
   NIP. 197902102002121001
2. **Dr. Wiranto M.Kom., M.Cs.** ( )
   NIP. 196612301993021001
3. **Hasan Dwi Cahyono S.Kom., M.Kom.** ( )
   NIP. 198205242014041001
4. **Sari Widya Sihwi S.Kom., M.T.I.** ( )
   NIP. 198304122009122003

Disahkan oleh

Ketua Program Studi Informatika,

Drs. Hanjang Harjito M.APP.Sc, Ph.D.

NIP. 196211301991031002

iv

HALAMAN PERSEMBAHAN

Tugas akhir ini ku persembahkan untuk

kedua orang tua dan adik adik ku tercinta,

teman teman informatika angkatan 2011,

keluarga besar UPT TIK UNS.

MOTTO

"Barang siapa yang menempuh jalan untuk mencari suatu ilmu. Niscaya Allah memudahkannya ke jalan menuju surga"
HR. Turmudzi

"Ya Allah aku memohon kepada-Mu ilmu yang bermanfaat, rizki yang baik, dan amal yang diterima"
HR. Ibnu Majah

"Maka apabila kamu telah selesai (dari sesuatu urusan), kerjakanlah dengan sungguh-sungguh (urusan) yang lain,"
Q.S. Al Insyirah: 8

"The quieter you become the more you can hear"
Baba Ram Dass

"The First Rule of Programming: It's Always Your Fault"
Coding Horror

KATA PENGANTAR

Segala puji penulis panjatkan kehadirat Allah atas limpahan nikmat, hidayah dan inayah-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul "Klasifikasi dan Pembangkitan Indeks Untuk Pencarian Koleksi Dokumen Online Dengan Menggunakan Metode Naive Bayes Classifier dan Vector Space Model".

Penulis menyadari bahwa tugas akhir ini masih jauh dari kesempurnaan, baik dari segi penulisan maupun materi. Walaupun demikian penulis berharap semoga tugas akhir ini dapat bermanfaat bagi berbagai pihak. Penulis mengucapkan terima kasih kepada semua pihak yang telah meluangkan waktu untuk memberikan bimbingan dan saran sehingga laporan ini dapat berwujud sebagaimana yang diharapkan, terutama kepada:

1. Ayah, Ibu dan segenap keluarga penulis yang telah memberikan kasih sayang, kesabaran, pengorbanan, do'a serta semangat kepada penulis.
2. Bapak Ristu Saptono, S.Si., M.T. dan bapak Dr. Wiranto M.Kom., M.Cs. selaku dosen pembimbing tugas akhir atas kebaikan dan bimbingan selama penyelesaiaan tugas akhir ini.
3. Para staff dan teman teman maganger UPT TIK UNS yang telah membantu banyak dalam penyelesaian tugas akhir ini.

Surakarta, Januari 2016

Penulis

KLASIFIKASI DAN PEMBANGKITAN INDEKS UNTUK PENCARIAN KOLEKSI DOKUMEN ONLINE DENGAN MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER DAN VECTOR SPACE MODEL

ARBA SASMOYO
Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Sebelas Maret

ABSTRAK

Universitas Sebelas Maret memiliki banyak repositori dokumen *online*. Mengelola repositori dengan jumlah banyak tidaklah mudah. Dengan banyaknya jumlah repositori dokumen tersebut justru mempersulit pengguna dalam mencari dokumen. Selain itu, metode pencarian pada beberapa repositori dokumen kurang optimal karena hanya mempertimbangkan judul saja.

Oleh karena itu, pada penelitian ini diajukan sebuah metode untuk mengindeks dan mencari dokumen yang tersebar di beberapa repositori. Terdapat beberapa langkah untuk mengindeks dokumen yang berbeda antara dokumen berbahasa satu dengan bahasa lain. *Naive Bayes Classifier* digunakan untuk mengklasifikan sebuah dokumen berdasarkan bahasanya. Selanjutnya, pencarian dokumen dilakukan menggunakan algoritma *Vector Space Model*. Proses klasifikasi dan pencarian diuji menggunakan perhitungan *accuracy*, *precision* dan *recall*. Hasilnya, *Naive Bayes Classifier* memiliki *accuracy* 97,62%, *precision* dokumen Indonesia dan Inggris 98,30% dan 95,56%, dan *recall* dokumen Indonesia dan Inggris 95,28% dan 98,17%. Sedangkan *Vector Space Model* memiliki *precision* dan *recall* sebesar 26,59% dan 100%.

Kata kunci: *Naive Bayes Classifier*, Pengindeksan dokumen, Repositori dokumen, *Vector Space Model.*

CLASSIFICATION AND INDEX GENERATION FOR SEARCHING ON ONLINE DOCUMENT REPOSITORIES USING NAIVE BAYES CLASSIFIER AND VECTOR SPACE MODEL

ARBA SASMOYO

Department of Informatics, Faculty of Mathematics and Natural Sciences, Sebelas Maret University

ABSTRACT

Sebelas Maret University has many online document repositories. Managing many document repositories is not a simple task. As the number of document repository increases, users will have difficulty searching for a document across multiple repositories. Poor searching method on document repository also give users evenmore bad experiences.

This research propose a method to index and search document which are located accross multiple document repositories. There are some steps to index documents, and some of them are languange specific. Naive Bayes Classifier will be used to classify document according to its language. Document searching will use Vector Space Model algorithm. Document classification and searching will be tested using accuracy, precision and recall. The results showed that Naive Bayes Classifier has accuracy 97.62%, precision for Indonesia and English 98,30 and 95.56% and recall for Indonesia and English 95,28% and 98,17%. Meanwhile Vector Space model has precision and recall 26,59% *and 100%.*

Keywords: *Document indexing, Document repository*, *Naive Bayes Classifier*, *Vector Space Model*.

DAFTAR ISI

DAFTAR TABEL

DAFTAR LAMPIRAN

DAFTAR GAMBAR